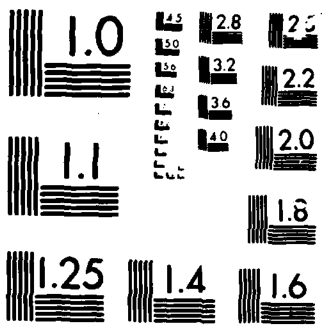
AD-A167 033 A NOTE ON ESTIMATING THE IMPROVEMENT IN STEIN-TYPE 1/1
ESTIMATORS(U) STANFORD UNIV CA DEPT OF STATISTICS
A E GELFAND 10 APR 86 TR-371 N00014-86-K-0156

UNCLASSIFIED F/G 12/1 NL



MICROCOP'   CHART

A NOTE ON ESTIMATING THE IMPROVEMENT IN STEIN-TYPE ESTIMATORS

BY

ALAN E. GELFAND

TECHNICAL REPORT NO. 371

APRIL 10, 1986

Prepared Under Contract
N00014-86-K-0156 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. INTRODUCTION

It is by now well-known that if $X_{p\times 1} \sim N(\theta, \sigma^2 I)$, $p \geq 3$, σ^2 known, X is an inadmissible estimator of θ under arbitrary positive definite quadratic loss. If σ^2 is unknown but an independent estimate of σ^2 is available, this conclusion still holds. An identity due to Stein (1973) has been instrumental in this development. A by-product of this identity is the fact that under a simple integrability condition a unique unbiased estimator of the risk of estimators of the form $X + f(X)$, $f_{p\times 1}$, can be provided. Efron and Morris (1976) discuss this issue in detail both with σ^2 known and unknown. They note that these risk estimators may be employed to select a best estimator from a class of estimators.

We address the following practical issue. If we select an estimator of θ known to improve upon X, how should we estimate this improvement? The above unbiased estimator need not be admissible under squared error loss; several attractive alternatives will be proposed.

2. ALTERNATIVE ESTIMATORS

Suppose $X \sim N(\theta, I)$. Let $Z = X^T X$ and consider the crude James-Stein estimator (pulled toward $\theta = 0$ w.l.o.g.),

$(1-(p-2)/Z)X$. Under loss $(\theta-a)^T(\theta-a)$, this estimator uniformly improves upon X with improvement

$$I = g(\lambda) = (p-2)^2\ E_\theta(Z^{-1})$$

where $\lambda = \theta^T\theta/2$. Hence $\hat{I}_1 = (p-2)^2\ Z^{-1}$ is immediately unique unbiased for $g(\lambda)$. Obviously $g(\lambda)$ decreases in λ from $g(0) = p-2$ (worthwhile improvement occurs when λ is small). Hence

$$\hat{I}_2 = \begin{cases} (p-2)^2\ Z^{-1}, & Z \geq p-2 \\ (p-2), & Z < p-2 \end{cases}$$

dominates $\hat{I}_1$. In fact, it is clear that $\hat{I}_1$ can be significantly improved upon when λ is small since in this case $\mathrm{var}(\hat{I}_1) \simeq 2(p-4)^{-1}(p-2)^2$ (the value at $\lambda = 0$) which is more than twice I regardless of p. Alternatives to $\hat{I}_2$ will now be developed using the fact that Z has a noncentral chi-square distribution, i.e. $Z \sim X_p^2(\lambda)$. $g(\lambda)$ is then a function of the noncentrality parameter with the explicit form

$$g(\lambda) = (p-2)^2 \sum_{\ell=0}^{\infty} \lambda^\ell e^{-\lambda}(p-2+2\ell)^{-1}/\ell! \qquad (2.1)$$

The UMVUE for λ is $(Z-p)/2$ suggesting the estimator

$$\hat{I}_3 = \begin{cases} g((Z-p)/2, & Z \geq p \\ p-2 & , Z < p \end{cases}.$$

The MLE for λ has been discussed by Meyer (1967). With a single multivariate normal distribution, it is clearly $Z/2$ whence $g(Z/2)$ is the MLE for $g(\lambda)$. $g(Z/2)$ is always less than $\hat{I}_3$ and when λ is small, regardless of p, this underestimation badly inflates mean square error relative to that of $\hat{I}_3$.

Improvements under squared error loss to the UMVUE for λ are discussed in Perlman and Rasmussen (1975) and in Neff and Strawderman (1976). One such estimator is $\hat{\lambda}^* = [(Z-p)/2 + (p-4)/Z]^+$, $p \geq 5$, suggesting the estimator $\hat{I}_4 = g(\hat{\lambda}^*)$. However, $I_4 \leq I_3$ and simulation (to be described shortly) shows that for λ small the $(p-4)/Z$ term drastically inflates the bias and variance of $\hat{I}_4$ relative to $\hat{I}_3$.

A more direct approach is to investigate Bayes estimators for $g(\lambda)$ under squared error loss. Suppose $\theta \sim N(0,\gamma I)$, i.e. $\lambda \sim \gamma/2\ \chi^2_p$ whence $\lambda | Z \sim a/2\ \chi^2_p(aZ/2)$ where $a = \gamma/(\gamma+1)$. We seek $E(g(\lambda)|Z)$. Using (2.1) we obtain

$$E(g(\lambda)|Z) = (p-2)^2 \sum_{\ell=0}^{\infty} (p-2+2\ell)^{-1}/\ell!\ \int \lambda^{\ell} e^{-\lambda} f(\lambda|Z)d\lambda.$$

Letting s = (p-2)/2 and performing the integration, we obtain

$$E(g(\lambda)|Z) = \frac{(p-2)^2}{2} \sum_{j=0}^{\infty} \left(\frac{aZ}{2}\right)^j \frac{e^{-aZ/2}}{j!}$$

$$\cdot \sum_{\ell=0}^{\infty} (s+\ell)^{-1} \binom{s+\ell+j}{\ell} \left(\frac{a}{a+1}\right)^{\ell} \left(\frac{1}{a+1}\right)^{s+j+1} . \qquad (2.2)$$

We may interpret (2.2) with J~Poisson(aZ/2), L|J ~ Negative Binomial ($\frac{a}{a+1}$, s+J+1) (where s need not be an integer) and

$$E(g(\lambda)|Z) = \frac{(p-2)^2}{2} E(E((s+L)^{-1}|J)). \qquad (2.3)$$

The identity

$$\sum_{\ell=0}^{\infty} (s+\ell)^{-1} \binom{s+\ell+j}{\ell} \left(\frac{a}{a+1}\right)^{\ell+s}$$

$$= \sum_{k=0}^{j} \binom{j}{k} (k+s)^{-1} a^{k+s}$$

(derivable by considering the indefinite integral with respect to a of $a^{s-1}(a+1)^j$ directly or through its equivalent negative binomial expansion) leads to

$$E(g(\lambda)|Z) = (p-2)^2/2(a+1) \sum_{j=0}^{\infty} \sum_{k=0}^{j} \frac{e^{-aZ/2}}{j!} \left(\frac{aZ}{a(a+1)}\right)^j \binom{j}{k} (k+s)^{-1} a^k .$$

Interchanging order of summation and summing over j, we obtain the Bayes rule in simple form:

$$E(g(\lambda)|Z) = (a+1)^{-1} g(a^2 Z/2(a+1)) . \qquad (2.4)$$

Letting $\gamma \to \infty$, i.e. $a \to 1$, in (2.4) leads to $\hat{I}_5 = 1/2\ g(Z/4)$. As noted in Perlman and Rasmussen, this resultant "noninformative" prior on θ yields a prior on λ which is nonuniform and, in fact, is biased against small λ. The fact that $\hat{I}_5$ is at most $(p-2)/2$ reflects this. A more appealing "empirical" Bayes estimator is obtained by estimating $(\gamma+1)^{-1}$ in (2.4) by pZ^{-1} (as in Perlman and Rasmussen) resulting in

$$\hat{I}_6 = \begin{cases} \dfrac{Z}{2Z-p}\, g\left(\dfrac{(Z-p)}{2} \cdot \dfrac{Z-p}{2Z-p}\right) , & Z \geq p \\ p-2 , & Z < p \end{cases} .$$

From (2.1), $g(\lambda) = (p-2)^2 E(p-2+2L)^{-1}$ where $L \sim \text{Poisson}(\lambda)$. Consider the following conditional estimation problem. Treating L as a parameter, estimate $\gamma(L) = (p-2)^2(p-2+2L)^{-1}$. This idea is also suggested by (2.3) and such an approach

was discussed by Stein (1964) in conjunction with the estimation of the variance of a normal distribution with unknown mean. Given L, $Z \sim \chi^2_{p+2L}$ and we seek estimators based on Z of $\gamma(L)$. In this setting $\hat{I}_1$ is again the UMVUE and again the truncation in $\hat{I}_2$ is appropriate. Since $E(Z-2|L) = p-2+2L$, the estimator

$$\hat{I}_7 = \begin{cases} (p-2)^2(Z-2)^{-1} & , \quad Z \geq p \\ p-2 & , \quad Z < p \end{cases}$$

may be considered. $\hat{I}_7 \geq \hat{I}_2$. For small λ simulation reveals $\hat{I}_7$ to be less biased with smaller variance than $\hat{I}_2$. For large λ, Z will likely be greater than p whence $\hat{I}_7 \approx (p-2)^2(Z-2)^{-1}$, $\hat{I}_2 \approx \hat{I}_1$. Now $\hat{I}_2$ is nearly unbiased and simulation reveals that $\hat{I}_2$ will have smaller variance than $\hat{I}_7$.

The MLE of L is not easily obtained, hence for γ as well. Under squared error loss, $(\gamma(L)-a)^2$, shrinkage of $\hat{I}_1$ is suggested, i.e. amongst estimators of the form $e\hat{I}_1$ the optimal e is $(p-2+2L)^{-1}(p-4+2L)$. Consider Bayes estimates of $\gamma(L)$ under this loss. Denoting the prior on L by π, the Bayes rule vs. π, $\delta_\pi(Z)$, becomes

$$\delta_\pi(Z) = \frac{(p-2)^2 \sum_{\ell=0}^{\infty} (p-2+2\ell)^{-1}(Z/2)^{\ell}\ \pi(\ell)/\Gamma(\frac{p+2\ell}{2})}{\sum_{\ell=0}^{\infty} (Z/2)^{\ell}\ \pi(\ell)/\Gamma(\frac{p+2\ell}{2})} .$$

If we define

$$h_\pi(Z) = \sum_{\ell=0}^{\infty} (p-2+2\ell)^{-1}(Z/2)^{(p-1+\ell)/2}\ \pi(\ell)/\Gamma(\frac{p+2\ell}{2}) \quad (2.5)$$

then straightforwardly,

$$\delta_\pi(Z) = (p-2)^2 h_\pi(Z)/2Zh_\pi'(Z) . \quad (2.6)$$

Clearly, $\hat{I}_1$ can't be Bayes vs. any prior, i.e. we would need $h_\pi(Z) = e^{Z/2}$, impossible by equating coefficients in (2.5). A shrinkage estimator arises if $h_\pi'/h_\pi > 1/2$. Two priors yielding simple expressions for $\delta_\pi(Z)$ are: (i) $\pi(\ell) = (\frac{p}{2} -1+\ell)\Gamma(\frac{p}{2} -1+\ell)/\ell!$ (mass on large ℓ) resulting in $\hat{I}_8 = (p-2)^2(p-2+Z)^{-1}$. $\hat{I}_8$ underestimates $g(\lambda)$ and when λ is small simulation reveals this to critically inflate mean squared error. (ii) $\pi(0) = (p-2)/p$, $\pi(1) = 2/p$ (mass on small ℓ) resulting in $\hat{I}_9 = (p-2)^2(p^2(p-2)+2pZ)^{-1}(p^2+2Z)$. With increasing λ the bias in I_9 tends to $(p-2)^2/p$ again critically inflating mean squared error. A uniform prior on L does not

yield an estimator in closed form. However, for small λ, this estimator tends to shrink $\hat{I}_1$; for large λ, it tends to expand $\hat{I}_1$.

A simulation based on 7000 replications at each λ and p was developed to examine the $\hat{I}_j$, $j = 1,\ldots,9$. Estimators involving g are most easily computed from (2.1) in recursive fashion with double precision, i.e. by writing $g(\cdot) = (p-2)^2 \exp(\cdot) \sum_{\ell=0}^{\infty} b_\ell$ where $b_{\ell+1} = \frac{p-2+2\ell}{p+2\ell} \cdot \frac{(\cdot)}{\ell+1} b_\ell$ and $b_0 = (p-2)^{-1}$. Table 1 presents an abbreviated summary for the best performers, I_2, I_3, I_6 and I_7.

3. EXTENSIONS

Extension to the case where $X \sim N(\theta, \sigma^2 I)$, σ^2 unknown, is immediate if an independent estimator $\hat{\sigma}^2$ of σ^2 is available based on a chi-square random variable with r degrees of freedom. The simple James-Stein estimator becomes $(1 - (p-2)r\hat{\sigma}^2/(r+2)Z)X$ which under squared error loss uniformly improves upon X with <u>relative</u> improvement

$$\frac{(p-2)^2}{p} \; \frac{r}{r+2} \sigma^2 E Z^{-1} = g(\lambda) \qquad (3.1)$$

where $\lambda = \theta^T\theta/2\sigma^2$. The independence of $\hat{\sigma}^2$ and Z enables straightforward development of estimators of $g(\lambda)$ paralleling

$\hat{I}_2 - \hat{I}_9$. The resulting estimators will depend on $\hat{\sigma}^2$ and Z only through $U = Z/\hat{\sigma}^2$. Details are omitted. In the context of estimation in a full rank linear model, let $Y = X\beta+\varepsilon$, $X_{n\times p}$, $r(X) = p \geq 3$, $\varepsilon \sim N(0,\sigma^2 I)$ and $\hat{\beta}_{OLS}$ be the ordinary least squares estimate of β. Then $T(\hat{\beta}_{OLS}) = (1-A)\hat{\beta}_{OLS}+A\beta^*$ where $A = c\hat{\sigma}^2/Q$, $\hat{\sigma}^2$ is the UMVUE of σ^2, $Q = (\hat{\beta}_{OLS}-\beta^*)^T X^T X(\hat{\beta}_{OLS}-\beta^*)$, $c = (p-2)(n-p)/(n-p+2)$ and β^* is a fixed vector, uniformly improves upon $\hat{\beta}_{OLS}$ under loss proportional to $(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)$ with <u>relative</u> improvement $\frac{p-2}{p} c\sigma^2 E(Q^{-1})$ analogous to (3.1). Estimators of this relative improvement will be expressible as functions of the "F-statistic," $(n-p)U/p$, or of $R^2(\beta^*)$, the sample multiple correlation coefficient resulting from fitting the adjusted regression model $Y-X\beta^* = X\beta+\varepsilon$.

In extending these ideas to minimax estimators of θ, other than the above James-Stein estimator, orthogonally invariant estimators of the form $(1 - (p-2)r\tau(U)/(r+2)U)X$ have been shown to uniformly improve upon X under squared error loss if, for example, $0 \leq \tau(\cdot) \leq 2$ and $\tau(\cdot)$ nondecreasing (Baranchik (1970)). The relative improvement of such estimators is shown to be (Efron and Morris)

$$\frac{p-2}{p} E \left\{ \frac{(p-2)r\ \tau(U)(2-\tau(U))}{(r+2)U} + 4\tau'(U)(1 + \frac{(p-2)}{r+2}\tau(U)) \right\}$$
$$= g(\lambda)$$

where λ again is $\theta^T\theta/2\sigma^2$. Without specification of τ it is unclear as to whether the implicit unbiased estimator of $g(\lambda)$ is admissible. Nonetheless, ideas of the previous section may be used to suggest alternative estimators.

1. A Comparison of Improvement Estimators

λ	p	$\varepsilon(\lambda)$	$E(\hat{I}_2)$	$MSE(\hat{I}_2)$	$E(\hat{I}_3)$	$MSE(\hat{I}_3)$	$E(\hat{I}_6)$	$MSE(\hat{I}_6)$	$E(\hat{I}_7)$	$MSE(\hat{I}_7)$
0.02	6	3.97	2.90	2.15	3.46	.92	3.48	.82	3.37	1.18
0.02	10	7.97	6.41	5.09	7.08	2.60	7.09	2.45	6.96	3.11
0.02	20	17.96	15.66	12.68	16.41	7.75	16.43	7.47	16.32	8.65
0.08	6	3.90	2.88	2.06	3.42	.91	3.45	.80	3.32	1.15
0.08	10	7.87	6.37	4.98	7.02	2.58	7.05	2.41	6.92	3.08
0.08	20	17.86	15.60	12.55	16.37	7.71	16.39	7.39	16.28	8.54
0.50	6	3.41	2.62	1.70	3.22	.86	3.26	.79	3.14	1.09
0.50	10	7.26	6.08	4.33	6.78	2.47	6.80	2.32	6.68	2.92
0.50	20	17.14	15.25	11.84	16.07	7.48	16.08	7.16	15.96	8.31
2.00	6	2.27	2.00	1.09	2.58	1.28	2.65	1.21	2.43	1.32
2.00	10	5.62	5.07	3.56	5.85	3.28	5.92	3.09	5.68	3.53
2.00	20	14.96	13.92	11.28	14.87	9.03	14.93	8.68	14.71	9.75
8.00	6	.88	.87	.21	1.11	.49	1.20	.55	1.01	.39
8.00	10	2.83	2.81	1.39	3.28	2.28	3.39	2.32	3.12	2.04
8.00	20	9.80	9.75	8.51	10.55	10.27	10.68	10.16	10.31	10.07
24.50	6	.31	.31	.01	.34	.01	.36	.02	.33	.01
24.50	10	1.16	1.15	.10	1.24	.15	1.31	.18	1.20	.13
24.50	20	4.95	4.94	1.41	5.21	1.81	5.34	1.96	5.09	1.64

REFERENCES

Baranchik, A. (1970), "A family of minimax estimates of the mean of a multivariate normal distribution," Annals Math. Statist., 41, 642-645.

Efron, B., and Morris, C. (1976), "Families of minimax estimators of the mean of a multivariate normal distribution," Annals of Statistics, 4, 11-21.

Neff, N., and Strawderman, W. (176), "Further remarks on Estimating the Noncentrality Parameter of a Noncentral χ^2 Distribution," Comm. in Statist., Theory and Methods A5, 1, 65-76.

Perlman, M., and Rasmussen, U. (1975), "Some Remarks on Estimating A Noncentrality Parameter," Comm. in Statist., 4, 455-468.

Stein, C. (1964), "Inadmissibility of the Usual Estimator for the Variance of a Normal Distribution with Unknown Mean," AISM, 16, 155-

_____ (1973), "Estimation of the Mean of a Multivariate Normal Distribution," Proc. Prague Symposium, Asymptotic Statist., 345-381.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 371	2. GOVT ACCESSION NO.: ADA 167033
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): A Note On Estimating The Improvement In Stein-Type Estimators	5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL REPORT
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Alan E. Gelfand	8. CONTRACT OR GRANT NUMBER(s): N00014-86-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, Stanford University, Stanford, CA 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Statistics & Probability Program Code 411SP	12. REPORT DATE: April 10, 1986
	13. NUMBER OF PAGES: 15
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multivariate normal, Stein-type estimators, Improvement, Non Centrality parameter.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In estimating the mean of a multivariate normal distribution under squared error loss, the MLE may be uniformly improved upon. For a dominating estimator, how may we estimate the improvement? Results from Efron and Morris enable a unique unbiased estimator of the improvement for orthogonally invariant estimators. However, this estimator needn't be admissible. We develop, for the James-Stein estimator, alternative improvement estimators.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

6-86

DTIC